DOCUMENT RESUME

ED 252 558 TM 850 027

AUTHOR McArthur, David; Cabello, Beverly
TITLE How to Use the DX SYSTEM of Diagnostic Testing. Methodology Project.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Nov 84
GRANT NIE-G-84-0112-P1
NOTE 50p.; For related document, see TM 850 026.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adaptive Testing; *Computer Assisted Testing; *Computer Software; Criterion Referenced Tests; *Diagnostic Tests; Elementary Secondary Education; Microcomputers; Student Evaluation; Teacher Made Tests; *Test Construction: Test Items
IDENTIFIERS Apple II; *Domain Referenced Tests; *DX Diagnostic Testing System

ABSTRACT

The DX SYSTEM of Diagnostic Testing is an easy-to-use computerized system for developing and administering diagnostic tests. A diagnostic test measures a student's mastery of a specific domain (skill or content area). It examines the necessary subskills hierarchically from the most to the least complex. The DX SYSTEM features tailored testing with immediate student feedback. In addition teachers can obtain an immediate student performance profile, plus additional information to help diagnose and remediate a student's problems. This manual provides detailed directions for operating the DX SYSTEM on an Apple II computer. The steps include: (1) "booting up"; (2) making a back-up copy; (3) the menu; (4) how to enter test items; (5) administering the test; and (6) how to obtain test results for individual students. In addition it describes six steps for developing diagnostic tests and test items: (1) developing a blueprint of the domain; (2) mapping prerequisite skills for the target domain; (3) developing a hierarchy or sequence of skills; (4) writing test items using common conventions of item writing; (5) reviewing the test items; and (6) field testing the items. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED252558

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE - NOVEMBER 1984

METHODOLOGY PROJECT

How To Use the DX SYSTEM of Diagnostic Testing

David McArthur

Project Director

Grant Number

NIE-G-84-0112, P1

Center for the Study of Evaluation

Graduate School of Education

University of California, Los Angeles

TM 850 027

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

How To Use the DX SYSTEM of Diagnostic Testing
Beverly Cabello

Table of Contents

	Page
Introduction	
DX, More than a Computerized Testing System.	1
The Diagnostic Testing System	
What the test measures	4
How to operate the DX SYSTEM	11
Computer orientation	12
Step 1 - Booting Up.	14
Step 2 - Making a Back-Up Copy	16
Step 3 - The Menu.	18
Step 4 - How to Enter the Test Items	19
Step 5 - Administering the Test.	23
Step 6 - How to Obtain the Test Results for Individual Students	24
How to develop the Diagnostic Test and Test Items.	26
Step I - Developing a Blueprint of the Domain.	29
Step II - Map of Prerequisite Skills for the Target Domain. .	33
Step II - Developing a Hierarchy or Sequence of Skills. . . .	34
Step IV - Writing Test Items Using Common Conventions of Item Writing	38
Step V - Review the Test Items	42
Step VI - Field Test the Items.	44
Troubleshooting: What to do If...	46

DX, More than a Computerized Testing System

The diagnostic testing system is, first and foremost, a system for diagnosing problems of individual students in a particular skill or content area. For convenience, we will call the target skill/content area a domain throughout the text. The critical features of this system are:

1. Each test is tailored to the individual student. Because the test is administered 'one on one', using a computer, the test taker's pattern of responses determines what kinds of items, and how many questions s/he will receive in the assessment of a particular domain.

2. Each student's performance within a domain is diagnosed by assessing the skill as a whole, and by assessing the prerequisite subskills, when necessary.

3. Each student receives immediate feedback after each response. Thus the testing event is also a learning occasion.

4. Because the test is computerized, the teacher can obtain a performance profile of each student immediately after the testing event, if the teacher so chooses.

5. Because of the structure of this test, the teacher can obtain several kinds of information which may help him/her to diagnose and remediate a student's problems.

6. The DX system is easy to use because the instructions for its operations are simple, and require no knowledge of specialized computer languages.

This manual will give you a step by step description of how to use the DX system for several purposes such as: writing your own test items; administering the test; obtaining test results. Before starting the operations section of the manual, however, we strongly urge you to read the next section, "A diagnostic model of testing," This section describes more fully how the system diagnoses students' strengths and weaknesses.

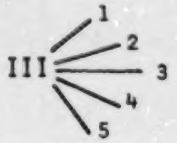
The Diagnostic Testing System

What the test measures

The goal of the diagnostic test is to measure individual students' mastery of a well specified domain. The assessment instrument examines the critical and prerequisite subskills or subconcepts which the student needs in order to master the domain. These skills should be arranged hierarchically from the most complex down to the least complex. If a student clearly has not mastered the domain, the system will administer items from domain subskills until it finds the student's maximum level of performance within the hierarchy. The testing system is designed so that the target domain and each of its prerequisite skills/concepts is measured by a cluster of five items. These item clusters are hierarchically arranged within the DX system such that each cluster represents a decision point for what item the test taker is subsequently given. (See figure 1) The hierarchy of clusters can be arranged in two to four levels. To progress along these levels, the students must make errors in two out of five items within a cluster. Mastery of a skill is indicated only if the student succeeds in four out of five items of the most complex skill cluster. The test stops when the test taker succeeds at a complete item cluster, or when s/he reaches the bottom of the hierarchy. At either point it is assumed that the individual has illustrated his or her degree of mastery with regard to the target skill or content area.

An additional feature of the cluster hierarchy is that the distractors for all items across all levels of the hierarchy fall within one to a maximum of four error categories. (See figure 2) These error categories

Figure 1
Sample Hierarchy of Items for the Domain of Inference

Cluster #V Target Domain: Making inferences from a passage	
Cluster #IV subskill: making inferences from two sentences	
Cluster # III subskill: integrating literal information from two sentences	
Cluster #II subskill: integrating and identifying information from one sentence	
Cluster #I subskill: identifying vocabulary and content from one sentence	

* 1 - 5 are individual question numbers.

Figure 2
Sample Distractor Categories for Inference Skill

Example 1

Distractor Categories

A: False but plausible information

B: Information that answers only part of the question

Cluster V

item 1	B	item 4	Correct
	A		B
	Correct		A
	B		B
	A		A
item 2	A	item 5	A
	B		B
	B		B
	A		A
	Correct		Correct
item 3	A		
	A		
	Correct		
	B		
	B		

Example 2

Distractor Categories

A: False but plausible information

B: Information that answers only part of the question

C: Information which is completely irrelevant to the question

Cluster V

item 1	A	item 4	C
	C		A
	A		Correct
	Correct		B
	B		A
item 2	C	item 5	C
	A		C
	B		A
	Correct		Correct
	A		B
item 3	A		
	Correct		
	B		
	B		
	C		

represent the teacher's hypotheses about the kinds of misunderstandings the test taker has with respect to the domain or the prerequisite skills.

An Example

Let's take a concrete example of how this test works (See figures 3 and 4). The domain we wish to assess is the student's ability to make inferences. The totality of this skill is measured by cluster #5, illustrated in figures 3 and 4. If the student errs in two out of the five items in this cluster, s/he will be presented with items from cluster #4, and so forth. Note that the student will be presented with items from a cluster until s/he makes two errors. Thus, an individual may successfully complete items 1-3, but err on items 4 and 5, then she will be presented with the next cluster. Or, the test taker may err on items 1 and 3, and then receive items from the next cluster.

Student #1 of the illustration (figure 3) successfully made inferences from small chunks of text (such as two sentences), but had difficulty integrating and making inferences from larger text chunks such as a paragraph. Student #2, (figure 4) on the other hand, had trouble identifying and integrating literal information from even one sentence. Yet we know that vocabulary (cluster #1) is not the problem.

At this point the user may want to hypothesize the reasons for a student's problem with the skill area. One helpful tool may be to look at the test taker's distractor patterns throughout the test. Notice that student #2 (figure 4) of the illustration seemed to consistently make the same type of error throughout the item clusters in the hierarchy. That is, s/he consistently selected the distractor falling in category A of our example: 'plausible but false information'. Thus one potential aspect of

Figure 3
How the Item Clusters and Response Categories Fit Together
and Example of #1 Test Taker's Path Through the Test

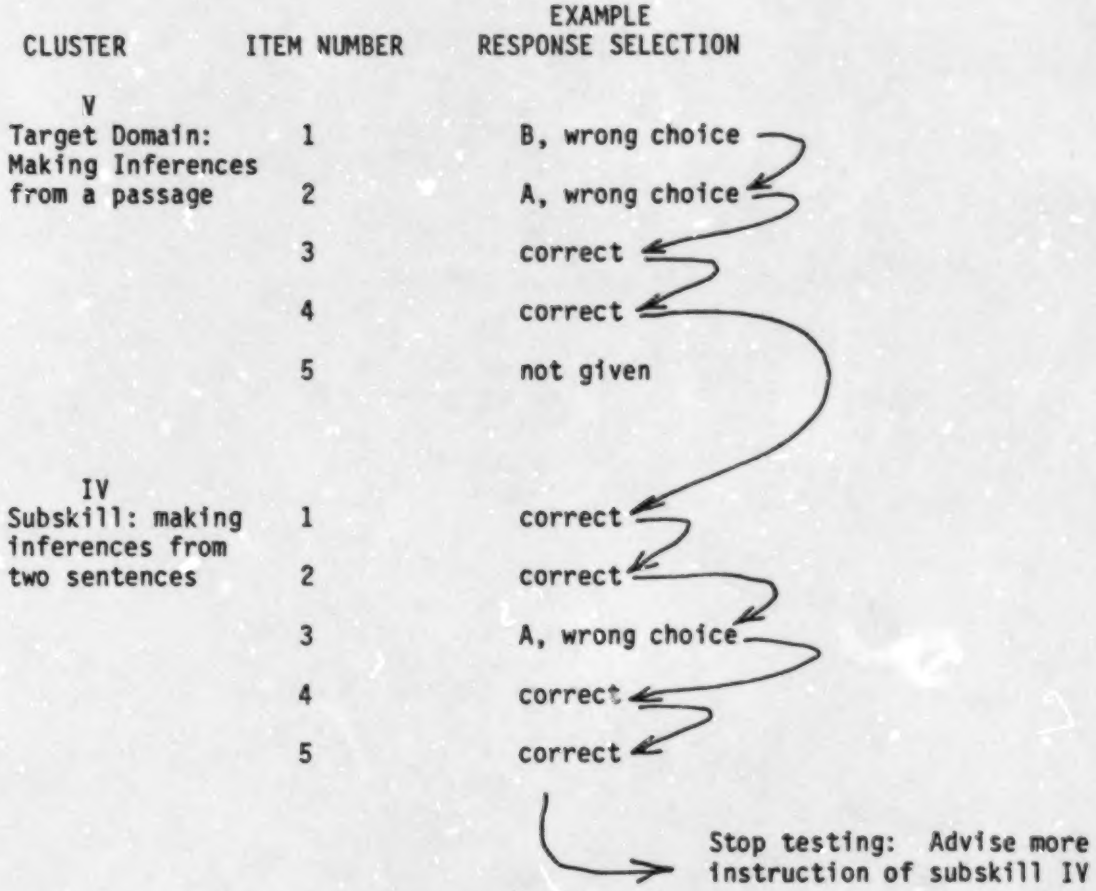
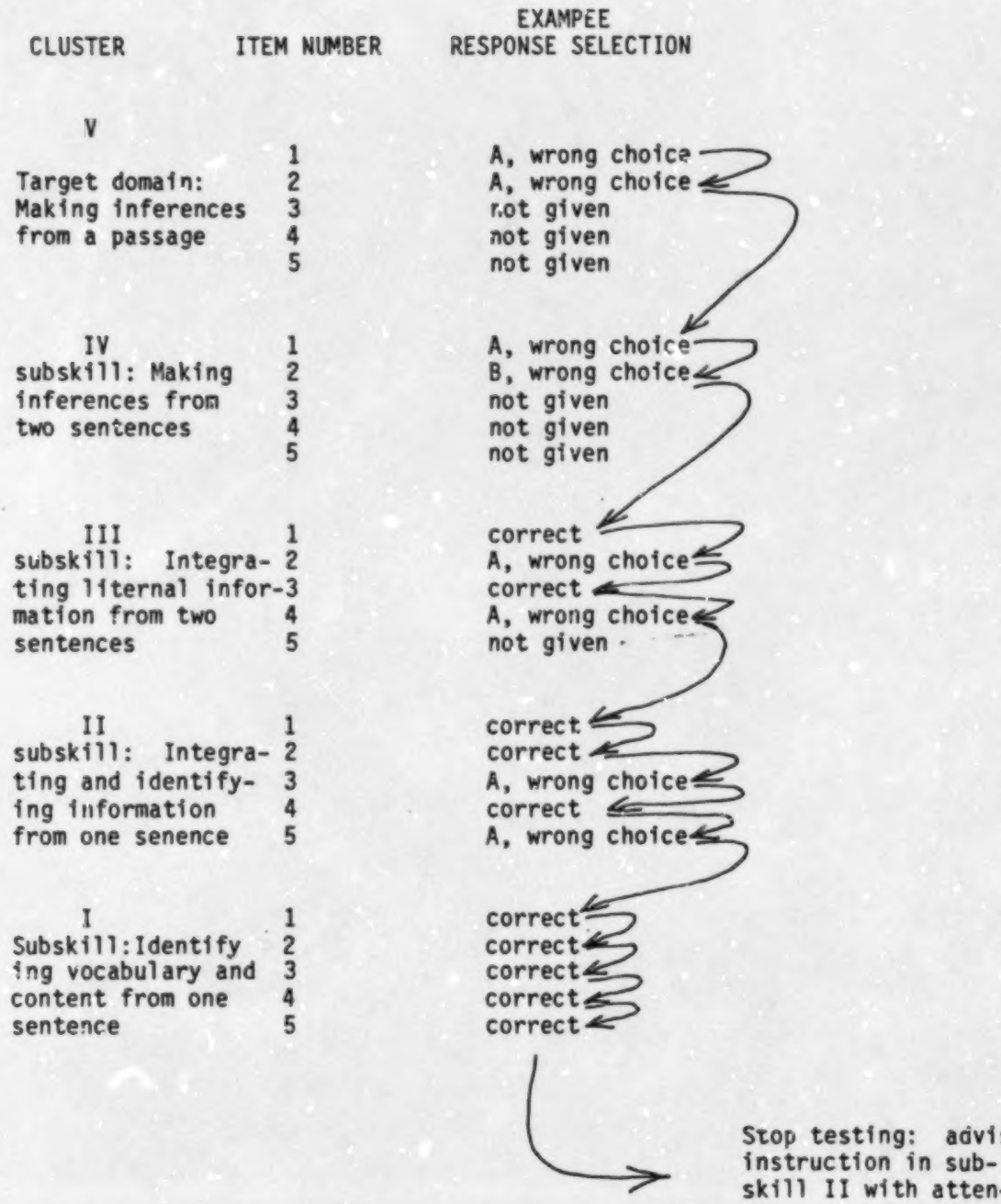


Figure 4
Example of #2 Test Taker's Path
Through the Test



instructional remediation may be to teach the student to distinguish plausible but false information from information which is clearly confirmed by the paragraph.

Student # 1 was successful in making inferences from two or more sentences. This student, however, seemed to select distractors which only partially answer the question (B). It's possible that the student is either misinterpreting the question (a test taking skill) or that s/he has not learned to integrate large chunks of text (an inference skill). In this case, the teacher may want to further test the student on this prerequisite skill.

HOW TO OPERATE THE
DX SYSTEM

The following pages will show you how to use the DX package. Just follow the steps described below. You will need the following equipment:

- an Apple II computer with a monitor (TV screen), two disk drives and a 48K memory
- the DX disk which includes some sample items
- blank disk
- a printer would be handy but is optional.

If you are already familiar with the use of the Apple II computer and the PASCAL language, make a back-up copy of the program disk and store it in a safe place. Then, go to Step 3, "The Menu" on page 14.

If you are not familiar with the use of an Apple II computer or with the Pascal language, be sure to read the following pages and follow all the steps very closely.

Computer orientation*

What do these computer components do?

The actual computer is the box on which the video screen is sitting. That box is filled with microchips which store information. The computer also brings information which has been programmed into it. This information tells the microcomputer how to interpret and store the information you and your students put into it.

---

* For a detailed description of how your computer and how it works please consult your Apple II manual.

The control card is simply a card containing microchips with special instructions on it. For example, you may want your screen to provide 80 columns of typing space instead of 40. The machine would need a particular control card inserted in order to do this. YOU DO NOT NEED TO MANIPULATE A CONTROL CARD. JUST MAKE SURE THAT THE COMPUTER HAS ONE BEFORE YOU START. The control cad should have been placed in the computer when it was set up.

The disk drive is like the starter on a car. You need it to start the computer. The "key" is the disk, in this case, you would need a DX disk. (The DX disk is also called a program disk.) The DX disk carries information which has been programmed into it as well as instruction for the interpretation and storage of new information. The disk "tells" the computer what to do. It also provides the information which you see on the screen. The disk also asks you what you want to do. Some computer come with 1 disk drive, others with 2 disk drives. You will see the importance of this in section 2 - "Making a back-up copy."

Step 1 - "Booting up"

When you insert the disk into the disk drive, you are "booting up" the disk. This is like putting the key into the starter and turning the car on. Here's how to do this:

a) Pick up the DX disk by holding it by the end with the label, labelled side up. Don't let your fingers touch the shiny plastic surfaces which show through the slots. The natural human oils could cause problems with the machine.

b) Insert the disk into the disk drive. Some drives may have a closing mechanism of some sort that you have to operate after inserting the disk.

c) Turn on the computer screen by using the switch on the rear left of your Apple II computer. Then turn on the video monitor as you would turn on a television.

d) At this point you should hear a whirring sound coming from the disk drive. Also, you may see a light turn on indicating that the computer is operating. At this point, a Menu will appear on the screen. You are now ready to begin the next phase of orientation.

A WORD OF CAUTION: Never remove or insert a disk into a drive that is still in the operating state. This would cause severe damage.

At this point, the manual will ask you to type certain things, press certain keys, and will refer to what appear on the key. For your convenience, all typing instructions will appear in the left hand column and all screen illustrations will appear on the right.

Step 2 - Making a Back-Up Copy

A "back-up" copy of the DX disk is simply an extra, emergency copy of the disk. You will use the copy to work from. You must store the original DX disk in a safe place, away from heat and moisture. It is important to make a copy of the original disk in case there is an unexpected problem with the system or the copy is damanged. Make the copy as follows:

1. Insert the DX disk into the #1 drive. Insert the copy disk into the #2 drive. Remember, hold the disk by the lable, with the lable side up.
2. Turn on the computer and the monitor. Wait for the whirring sound and the operating light.

To make a backup copy:

YOU TYPE	SCREEN SHOWS
	Hi, Welcome to the DX system. What would you like to do?
COPY (press return key)	

Before going any further, store the original disk, (the program disk supplied with this manual). From now on, you will use the back-up copy you just made.

Now that you have your back-up copy, you are ready to see what the DX system is all about. Place your DX disk in Drive 1.

Step 3 - The Menu
To begin by getting the menu:

YOU TYPE	SCREEN SHOWS
	Hi! Welcome to the DX system. What would you like to do?
SEE MENU (press return key)	MENU W Write new test items R Revise existing items A Administer the test G Get test results S Directory of sudents D Directory of test items Q Quit

(Type in letter indicating what you would like to do. See subsequent sections under the Menu headings for specific instructions).

* Press the return key each time you type in a command.

Step 4 - How to Enter the Test Items

This section will illustrate how to enter the test items, including the passage, item stem, questions, response choices and answer key. For information about how to develop effective test items, please see section _____.

There are a few important items to keep in mind as you enter the test items. First, remember that you have a limited space for entering the item passage, questions, response choices and answer key. This space consists of 80 columns (or spaces) across and 24 lines down the screen. The column space includes space for punctuation marks and spacing between words. Second, you have to label each passage and question set with a cluster label. These cluster labels will enable the computer to select particular items and questions for analysis and will help you to sort and access this information when you need it.

W Entering New Test Items

YOU TYPE	SCREEN SHOWS
W	What is the name of this skill cluster?
(Give a skill cluster name) OVERALL INFERENCE	How many questions will you give this skill cluste.
(Indicate number of questions) 2	What are the distractor categories?
(Give a name for each distractor category) A = FALSE BUT PLAUSIBLE INFORMATION B = INFORMATION THAT ANSWERS ONLY PART OF THE INFORMATION	
	Begin writing the text passage here. End with an asterisk.*
(An asterisk lets the computer know that you have finished entering the test passage.)	Many creatures of the world change completely as they grow from babies to adults. This complete change of form is called a metamorphosis. Baby frogs are a good example of metamorphosis. Adult frogs lay eggs. These eggs change into small fish-like creatures called polliwogs. They have no limbs. Their bodies change and they become frogs. When the frogs lay eggs, the process will begin again.* (Example test passage) Write question 1 here, including 4 answer choices. End with an asterisk * followed by correspondence between choice and distractor categories. Example: *1=A, 2=B, 3 = correct, 4=C

YOU TYPE	SCREEN SHOWS
(Type in question and answer choices) THE MAIN IDEA OF THIS PASSAGE IS: 1 Many creatures lay eggs. 2 Many creatures become fish as they grow. 3 Many polliwogs lay eggs. 4 Many creatures undergo metamorphosis as they grow. *1=A, 2=A, 3-B, 4=correct.	Write question 2 here, including 4 answer choices. End with an asterisk * followed by correspondence between choice and distractor categories.
(Type in question)	This passage is mostly about: 1. what polliwogs look like 2. how frogs grow from eggs 3. what fish look like 4. how polliwogs grow legs *1=a, 2=correct, 3=A, 4=B End of skill cluster, "Overall Inference". We are saving your test and questions on disk. Now write the text for the next question. Remember that you have 3 questions left.

R REVISE EXISTING TEST ITEMS

YOU TYPE	SCREEN SHOWS
E	Which test cluster would you like to edit?
(enter cluster name) Overall Inferences	
(To retype, delete, or correct words and puntuation marks, use the arrow keys and type in your corrections. To move to the next line, press the RETURN key.)	Test Passage 1 Many creatures of the world change completely as they grow from babies to adults. This complete change of form is called a metamorphosis. Baby frogs are a good example of metamorphosis. Adult frogs lay eggs. These eggs change into small fish-like creatures called polliwogs. They have no limbs. Their bodies change and they become frogs. When the frogs lay eggs, the process will begin again.* Many creatures lay legs. Many creatues lay (use arrow key and correct) Many creatures lay eggs.
(Now type any of the following:) U (to finish and update* the disk file with the latest changes) R (to return to revising this cluster again) E (to exit** without perserving any changes.)	

---

* To "update" means to permanently replace, correct or change what is on the disk. By permanent we mean that the next time you use the disk, the changes you have made will appear. If you do not want to preserve the changes, then press E for exit.

** By exit, we mean that you do not want to continue making changes in this particular cluster. You may, at this point, make changes in another cluster, or exit the program altogether.

Step 5 - Administering the Test

YOU TYPE	SCREEN SHOWS
	Welcome! Please enter your first and last name.
(first and last name, eg:) A	
Bonnie Brooks	Hi Bonnie. I will show you a passage and a question about it. Each question needs just one answer. When you read the question, give your answer by pressing A or B or C or D.
(type in your response)	Is A your choice?
A	Y/N
	Now type S (for start)
	(test begins)

The name the student types will be used by the computer as the label under which his/her test results will be stored. Whenever you try to obtain these particular results, you must give the computer the name just as the student typed it in -- including any spaces and punctuation marks. The test taker could have given any name, such as "banana" and the computer would accept it. It does not really matter how a test taker enters his or her name, so long as you are aware of the name being used. Later, you can use the same name for access to that student's results. So, if an examinee types in gibberish, then you may have trouble in gaining access to that student's file.

If the test taker accidentally or intentionally presses any other key on the keyboard other than A, B, C, or D, the "What is your choice?" question will be reprinted, including the four alternative key choices.

Step 6 - How to Obtain the Test Results for Individual Students

While the student has been taking the test, the computer has been recording the response patterns; and preparing diagnostic interpretations. In our example test the computer recorded the following

The number of correct and incorrect response at each cluster level.

A total for all of the clusters presented to the test taker.

The number of times the test taker selected distractors from each distractor category.

To obtain and read the test results for each student, you must return to the Menu. After the farewell greeting which signalled the end of the test, the Apple II cursor appeared at the bottom of the screen. Now

YOU TYPE	SCREEN SHOWS
	RUN MENU
(pause) Press return and wait	RUN MENU
G	Whose results do you want?

The results are stored as a file named for the person tested. To obtain a particular file, do the following:

(Type Test Takers Name)
Bonnie Brooks

SAMPLE TEST 1

Cluster	Items Correct	Wrong	Distractors
V	3	2	AB
IV	3	2	BB
III	4	1	B

Advice: Bonnie needs more work at clusters IV and V. Distractors suggest she needs to pay attention to B(Describe B hypothesis)

Do you want printed copy?

(Y?N)

The computer wants to know whether you want the test results printed. If you say N (no), the screen will just show you the file. If you say Y (yes), both the screen and the printer will give you the file. Let's assume that you want a printed copy:

Y ...Make sure your printer is ready.

To make sure that you are ready to print, check the following:

1. Is your printer connected to the computer?
2. Have you switched the printer on
3. Have you switched on the "on line" button located on front surface of the printer?
4. Does the printer have sufficient paper and ribbon in it?

HOW TO DEVELOP THE DIAGNOSTIC TEST AND TEST ITEMS

How to Develop the Diagnostic Test and Test Items

The following section is intended to provide guidelines for the development of a good diagnostic test structure and good diagnostic test items. It is by no means the only way of developing tests nor does it represent an overview of test development methodologies.

How does one go about building a good diagnostic test? The following five steps can guide this process:

I. Develop a blueprint of the skill or content area (to be labeled 'domain' for the remainder of this section) you want to diagnose. For example, if you want to develop a test for reading comprehension, what are the major skills that lead to good reading comprehension? How can you tell whether someone is a good comprehender? Specify this as clearly as possible.

II. Develop a map which describes the tasks and subtasks that are prerequisite to the target domain. For example, a prerequisite skill which allows readers to develop comprehension skills is the ability to decode chunks of text rapidly and accurately. Specify as many such important prerequisite skills as possible.

III. Develop a hierarchy or a sequence of skills going from the prerequiste skills or content for the domain, up through the critical skills of the domain. This will help you determine in developing hypotheses about the student's overall strengths and weaknesses by their performance on groups of items. It will also be a critical component in examining the test takers' error paterns.

IV. Write the test items based on the blueprint and prerequisite skill maps using common conventions of item writing.

V Review the test items to confirm their match to the item specifications and to assure that items do not contain extraneous complexities, unintended cues, or other technical flaws. Ideally, persons other than the test developer would conduct this review with a 'cold eye'.

VI. Field test the items to determine where item revisions are necessary, to clarify or revise the relationships - between the critical skills and the prerequisite skills, and to determine the number of items required for testing.

Each of these steps is described in the following sections.

STEP I: DEVELOPING A BLUEPRINT OF THE DOMAIN

This step involves identifying and specifying the domain you wish to diagnose; including what its critical skills and concepts (or knowledge base) are. The following questions serve as guides in formulating the target domain:

1. How much instructional time does it take to teach the domain? You'll want to select a domain which covers a reasonable amount of instructional time. For example, if a science curriculum consists of four major areas (i.e. biology, chemistry, physics, astronomy) and each of these is in turn divided into several subcomponents (i.e. biology: botany, ecology, etc.) it may be more efficient to develop a diagnostic test for each of the four major areas, but not for the area subcomponents, or for all four areas at once.

2. How does the objective relate to other higher-order skills? Be sure that the knowledge and skills you are testing and teaching reflect or are prerequisite to higher level thinking and problem solving skills. If we take the science example from above, the curriculum probably stresses the understanding and application of the scientific model for problem solving. Your test should examine the students' knowledge of this higher level skill. To examine factual knowledge without integrating it with thinking skills would result in an trivial exercise, not conducive to the betterment of instructional intervention.

3. How does the domain relate to long term curricular goals? As mentioned previously, your target domain should reflect an important part of your curriculum. However, all curricula have both short and long term goals. In our example science curriculum, the four areas and their subcomponents constitute definite time slots of instruction, units within the texts, etc. Yet the skill of applying the science model for problem solving is an overarching skill which crops up throughout the entire science curriculum. Thus you could develop a diagnostic science tests which holds this thinking skill as its target skill or domain, but which determines to what extent a student can use this skill across or within the four science areas.

To specify the domain then, you need to identify a critical part of the curriculum or a critical overarching skill or concept that would be fruitful to assess. One approach to doing this is to parse the entire semester's continuum for a content area (i.e. science or reading) into testable units. For example, you could take the hypothetical science continuum described above and examine each of the four areas, making sure to test students' progress using the scientific model of problem solving. Using this approach, you might see how easily students apply the model within and across the four major content areas.

Another important consideration is the level of cognitive complexity students are expected to function. For example, following Bloom (1956), does the skill of interest involve recall, application, analysis, synthesis? Or, following Gagne (197), does the skill represent concept learning (concrete or abstract), principles, procedures, problem-solving?

After completing step one, you should have a clear vision of the domain you wish to test, including its main components and boundaries. Let's use as an example the area of reading comprehension. Figure 5 illustrates a portion of a sample continuum for reading comprehension. Notice that the content of the passage is parsed into: science, social studies, math, reading and life skills. There are also some core skills which are emphasized throughout the units such as: inferring the main idea and identifying supporting details. Using such a continuum, our hypothetical test developers could specify their blueprint domain as follows:

Domain: The learner will analyze a given narrative or factual passage and correctly infer the following aspects;

1. main idea
2. concepts or relationsips which support the main idea.

Figure 5
Sample Reading Continuum

Time Frame	Unit	Comprehension Skill	Content Area/Life Skill	Decoding		Vocabulary
Sept./ Oct.	1	Main Idea Comparison Supporting Details Referents Other Details	Social Studies --- Narrative: fables, moral tales --- Resources: Government Information Pamphlets	Suffixes	-ive, -ical, -al	Word identification Vocabulary development (Synonyms/antonyms) (multiple meanings)
Oct./ Nov.	2	Character Main Idea Supporting Details Comparison	Science Reading --- Narrative: Biographic Styles Resources: Emergency Skill Pamphlets	Prefixes Suffixes	anti-, pro- -tiy, -ure	Word identification Vocabulary development (classification) (synonyms/antonyms)
Nov./ Jan.	3	Sequence Main Idea Supporting Details Character	Math Reading Narrative: Biographic Styles Resources: Assembly or Operating Instructions	Prefixes Roots	inter-, intra-, extra- rupt, duct, mit	Word identification Vocabulary development (multiple meanings) (classification)
Jan./ Feb.	4	Referents Cause/Effect Sequence Main Idea Supporting Details	Science Reading Science Fiction Resources: Application Directions	Roots Combining Forms	dict, ject, tract Centi, kilo	Word identification Vocabulary development (multiple meanings) (context clues)

STEP II: MAP OF PREREQUISITE SKILLS FOR THE TARGET DOMAIN

Notice that the continuum also indicates a number of prerequisite skills for the domain described previously. These could be specified as follows:

Prerequisite skills: In order to successfully master the domain or inference, the learner must have the following prerequisite skills:

1. Speedy and accurate decoding skills. In addition to accurately decoding words and groups of words, the reader must also recognize word parts which appear frequently and which indicate meaning such as suffixes, prefixes and roots.

2. The reader must have adequate knowledge of passage vocabulary and skills for identifying vocabulary meaning from context clues. (See Figure 6)

STEP III: DEVELOPING A HIERARCHY OR SEQUENCE OF SKILLS

Now that our hypothetical test development team has blueprinted the domain and mapped out the prerequisites, they need to plan either a hierarchy or a sequence of them. Why? There are several reasons for this. First, the purpose if a diagnostic test is not simply to identify where a student's problems are. A really useful diagnostic tool should provide some insight as to the kinds of misunderstandings or skill gaps a student has. By building a test structure around a skill hierarchy or sequence, the test developer can develop better hypotheses about the kinds of problems students may have. This would lead to better instructional remediation. A second reason for specifying the skill hierarchy is that it should guide the computer's test administration. Recall that this test is administered one on one by the computer. The computer must decide what item(s) to present the test taker based on the test taker's response. If the test taker starts the test with items from the most complex portion of the hierarchy--and fails--then s/he needs to be presented with items from less complex skill areas and so forth, until s/he reaches his or her maximum level of expertise. A student could indeed reach the "bottom" of the hierarchy. The test developer needs to provide the computer with a roadmap for doing this. Figures 6 and 7 which illustrate the domain blueprint and prerequisite skill map also show a suggested sequence of test items from each of the cells.

EXAMPLE:

Supose the teacher wants to diagnose students' inference skills for both narrative and scientific passages. A student could progress through the domain hierarchy as follows:

1a, 1b
2a, 2b
3a, 3c

If the teacher is interested only in diagnosing performance with one type of text (ie. narrative) then the path could be: 1a, 2a, 3a and so forth.

Figure 6
Example Prerequisite Skill Map

		Narrative Passages	Science Passages
decoding	prefixes		
	suffixes		
	roots		
	whole words		
vocabulary	recognition		
	context cues		

Figure 7
Example Domain Blueprint

Main Idea	1a	1b
Supporting concept or relationships	2a	2b
Details which strengthen the main idea or support	3a	3b

STEP IV - WRITING TEST ITEMS USING COMMON CONVENTIONS OF ITEM WRITING

There are a variety of formats that can be followed for item writing. The one described below combines models suggested by Popham (1980) and by Baker (1976) and includes the following components:

> General description - a brief description of the objective, skill, or knowledge to be measured.
>
> Sample item - a model of what test items are to look like, including directions to be given to students.
>
> Content limits - a description of the nature of the question that is to be presented to students.
>
> Response limits - a description of the response options provided to students for selected response items or for constructed response items the set of rules or criteria that are to be used to judge the quality of a students' response.

The first two components are relatively straightforward: a statement of the objective selected for testing and instruction and the sample item that has been devised for assessing it. Included here also the directions that will be given to students. Explicit attention to the directions early on helps to assure that they will be clear and that students will understand how to complete each item. Specifying content and response limits is more complicated and is explained in greater detail below.

Content limits describe the range of eligible content from which test items may be written. They may include rules for creating questions, rules for inclusion of prompts, cues, or additional materials (e.g., pictures, graphs, reading selections).

Content limits for selected response items define and restrict the characteristics, format, and eligible content and rules to be included in the item stem. By systematically including the different situations, contexts in which the skills are to be applied and/or the rules which define the assessed skill, test items can provide valuable diagnostic information, e.g., in what situations are students able to demonstrate a particular skill, what rules have students mastered? For instance, for a multiple choice item assessing students' skill in using appropriate pronouns, the content limits might be as follows: The item will present the student with a short (3 - 5 sentence) paragraph describing an action or event involving two or more named individuals.

> A blank will replace the named individual(s) in one sentence.
>
> Students will be asked to identify the pronoun which correctly completes the sentence.

Items will be written to exemplify the following rules:

> When the pronoun is the subject of a sentence or clause it should be in the nominative case.
>
> When the pronoun is the indirect object it should be in the objective case.

Such a systematic plan enables a test to diagnose which rule(s) are causing students' difficulty; the problem ascertaining the number of items to be written to reflect each rule is addressed in a late section.

Response limits provide rules for generating the correct response and incorrect alternatives for selected response questions and rules and criteria for judging the quality or correctness of a student's constructed response. Like the content limits, response limits help define the range

of eligible content but here the focus is on student responses: what discriminations are expected and reasonable; what are the characteristics of an acceptable response. For selected response items, response limits provide rules for constructing the correct answer and the distractors, or wrong answer alternatives for each item. These rules should assure distractors that represent common student errors and which thus may provide important diagnostic information. For example, response limits for the pronoun example described above might be as follows:

> Four alternatives will be provided for each item, the correct answer and three alternatives.
>
> The correct response will exemplify the proper application of the given rules and will reflect the appropriate gender and number.

Distractors will consist of the following:

- a pronoun in the correct case, but incorrect in number or gender;
- a pronoun in the incorrect case, but correct in number and gender;
- a pronoun representing an incorrect referent, but correct in case, number, and gender;

With such a set of alternatives, a student's wrong answer choice might provide information on whether he/she was having difficulty in identifying referents, was confused about case rules, and/or was having difficulties associated with number or gender.

The result of this specification process is a map for developing test items and likewise a map guiding instruction. Not only does the specification provide rules for developing multiple parallel test items, it

likewise can be used to generate relevant exercises for classroom practice, practice which will help students to acquire the specific skills they are intended to learn. So, although the process takes some time and effort, there are pay-offs.

STEP V - REVIEW THE TEST ITEMS

Once items are developed, the next step is to conduct a thorough review, considering two basic questions:

Do the items match their specification?

Are they free from technical flaws, i.e., do they follow conventional rules of item construction?

Do the items match their specification? The answer to this question is critical to establishing the content validity of the test. The process is straightforward: have each item examined by a colleague to compare its match with each element in the domain specification. That is, the description of eligible subject matter and item features provided in the content limits need to be compared with the content and features of the test question; the specification rules for creating correct and incorrect answer alternatives are compared with the actual set provided in selected response items. The items are checked also to see that they follow the prescribed format and that appropriate directions are given. While covered again under "technical flaws," check also to assure that the language used in the items is not unnecessarily difficult or complex and they are free from content that might be biased against particular groups of students. Where any problems are encountered, suitable item revisions need to be made. In some case there may be instances where the specification needs to be changed instead.

Are the items free from technical flaws? The review process here is also straightforward. Check the items against the general rules for constructing test items of particular types, and where flaws are detected, correct them. As with the content review described above, it is

preferable to have the review conducted by a colleague (the advantage of having a "cold," objective eye).

STEP VI - FIELD TEST THE ITEMS

Field testing the items is a final step in the test development process to assure high quality items, to verify the test structure, and to determine the number of items that will be needed to reliably diagnose students' performance. The optimal field test procedures involve a two stage process: 1. pilot test the items with a small sample of students to check their appropriateness; 2. administer the test to a larger sample to validate the subskills that need to be included in the test and the number of items required for each skill and subskill.

The purpose of the first pilot test is to determine whether the items are appropriate for students and to identify items that are potentially in need of revision. Have a small number of students (15-20) who are similar to the intended student popluation take the entire test and provide feedback on any problems they encounter, e.g., vocabulary or directions that are unclear; items where there seem to be more than one (or no) right answers. This feedback helps indicate where revisions are necessary.

Item difficulty indices (the percent of students who answer an item correctly) also help signal potential problem items. Because they are based on the same blueprint, one expects similar item difficulties for all items measuring the same subskill or task. For example, item difficulties for four of the five items measuring a particular subskill are .5-.7; however the difficulty of the fifth item is .25. This latter item should be re-examined to determine whether it is aberrant and is unintentionally confusing the correct response, whether it matches the specification, whether it represents a problem type that is different from the other items, whether the correct answer has been miskeyed and/or whether there

are typographical or other errors in the items. Any detected errors or deviations will need to be corrected. Item difficulties can also be used to help judge the appropriateness of the test for particular students. In order to Be useful in a diagnostic sense, a test should measure target skills which are diffcult for a substantial number of students: If all or most students get all or most of the items correct, there is little to diagnose.

Troubleshooting:	What To Do If.....
You type in a response to a computer question or you punch a command key such as RETURN and nothing happens.	1. Make sure that you type in response exactly as they are shown in this manual. If you add an extra letter or puntuation mark, or leave one out, the computer will not understand you. 2. If you press a response key and there is a "beep" sound as well as a screen change, then the computer has not understood or accepted whather you typed in. Just repunch the key until the computer responds with a beep.
You get an error message like:	1. Press the RETURN key and retype your command a second time. Be extra careful of spelling, spacing and punctuation. Then press RETURN.
You respond to an error message and the computer still won't work!...	Take a deep breath and: 1) turn the computer off by using the switch on the rear left side; 2) wait a few seconds; 3) turn it back on. The Menu will automatically reappear.
You can't find a file, or the computer says: FILE NOT FOUND	Make sure that you have not made a typographical error on the file name. Retype the correct file name (don't forget the DX) and try again. If that doesn't work, you may have the wrong file name or the student may have used a file name that you were not expecting. So, you must examine the list of file names: